NIEHS Guide to Training Evaluation

1. Learning, Sharing and Being Accountable

While there are many aspects of our programs that we understand well, often how and why things happen the way they do in a program can be a mystery. We frequently operate with limited time and resources. The ability to reflect and learn is at a premium. Implementation evaluation is a means to ensure that we create the time to reflect, that we gather information in a systematic way so that we can learn from those reflections. The products of implementation evaluations are for sharing what is learned with others and for being accountable. Without the systematic process of evaluation, over time, much of what is learned in carrying out a program can be forgotten. If we are to build our capacity to effectively conduct health and safety education programs, as well as build a broader health and safety movement, it will require that we:

- **learn** by organized reflection on information collected in an evaluation plan,
- **share** the lessons we learn, and
- **are accountable** to students, staff, funders, and the organizations that sponsor our programs.

A program history documented in an implementation evaluation can serve these purposes.

We have defined evaluation methods broadly because we believe there are many underlying issues that affect how evaluations will be conducted and what methods will be used. We wanted to bring some of these to the surface so that they can be considered head on.

Planning an evaluation is not a straight forward, step by step process. There are decisions to make each step along the way about:

- whose interests to serve in the evaluation
- who to involve in the planning and carrying out the evaluation
- whether to use methods that focus on information that tells how much of something happened (quantitative methods) or how, what and why things happened (qualitative methods), or to use both
- whether to conduct an evaluation that targets implementation of the program or one that targets impacts or to do some of both
- whether or not to use goals and objectives or some other factors as the basis of the evaluation

Each decision made will affect many other decisions. There's no perfect starting place and there are a number of routes you can take to develop an effective evaluation plan.

Two Types of Boxes

Jargon. While we have tried to limit the evaluation jargon in this section we wanted to acquaint you with some of the terms that you may encounter. We have placed these terms in "Jargon Boxes" so that you could easily identify them. Once introduced we've used some jargon just because it saves words and space.

Work. While this chapter is not a workbook, there are specific places where "Work Boxes" are provided for you to write and work through key aspects of the evaluation plan. In some cases we use sample Work Boxes already filled out. In this case blank boxes are provided at the end of the chapter.

2. Evaluation: In Whose Interests?

Real World Choices

The real world requires making sometimes difficult and complex choices. Time and resources for evaluation are usually very limited. Making these choices requires examination of a number of issues including:

- **Points of View** - What are the various points of view of the interested parties?
- **Priorities** - Whose interests are most important to be served at this time?
- **Accommodation** - How can the evaluation be best designed so that a number of interests are satisfied?
- **Ownership and Value** - How can the evaluation be designed so that those most closely involved with the program, students, instructors and other staff, see it as useful and valuable and support its aims and implementation?

These questions about the evaluation need to be addressed from the beginning.

Points of View

Every education program and every evaluation has a point of view. It emphasizes some interests while downplaying or ignoring others. Often the overwhelming factor in decisions about the purpose and scope of an evaluation is the need to satisfy the funding source. While it is only reasonable that the interests of the funding source occupy a major position in decisions about an evaluation, focusing exclusively on this interest may actually serve to undermine support for evaluations among others involved in the program. With such an exclusive focus on funders, instructors, staff, or students may come to see evaluation as something *done to them not for them*. When evaluators place all their focus on satisfying one interest, like the funding agency, they may not notice that instructors and students grow to resent the evaluations. When asked in a recent needs assessment conducted for this manual how instructors felt about evaluations conducted for their programs, one instructor summed up his frustration:

> ***It's directly dependent on where the evaluation came from. Those by an outsider are really looked down on. A person with no knowledge of your program, your political situation - a "cold fish evaluation." They [instructors] say, "how long will it take to get this [person] out of here"? They're going to stroke them - get through this evaluation and then go back to doing it the way***

> ***they want to. The evaluation is not worth a tinkers damn. When evaluators give good useful feedback good instructors put it to use. [You] build mutual respect. [When it's] simply a buck and a truck, they'll stroke em and get through it.***

Another instructor interviewed for the needs assessment commented:

> ***We're not involved at all, but we should be. We get a report or have a meeting ..... Instructors ought to be involved from the beginning and should help design the evaluation of courses they deliver.***

These instructors' feelings we're echoed by a program administrator who said:

> ***They [instructors and students] have not been involved in any real way except [the end of program evaluation sheets] .... They're not actually involved at all, but they ought to be.***

Evaluation, like other aspects of worker education programs, has the potential to either foster or diminish the sense of ownership among those participating in the program, a sense that is crucial to a program's success. The potential of a program evaluation to provide useful information to help programs improve, grow and prosper is perhaps equaled by its potential to foster alienation among key program constituents. None of those who contributed to the needs assessment survey were totally cynical about program evaluation. They simply recognized that without the input of a variety of interests the evaluations could not fully serve their programs.

Funding agencies, policy makers, program administrators, evaluators, instructors, other staff, and especially the worker -students, each have a potentially unique point of view and interest in an evaluation. Some of these interests may compliment each other, while others may conflict. Funding agencies, program administrators and evaluators need to recognize that it may be necessary for evaluations to address a number of these interests at the same time. The following chart is meant to represent the voices of a number of interested parties concerning the most important issues for program evaluation.

Evaluation Points of View				
Evaluation Type	**Workers/ students**	**Instructors**	**Program Administrators**	**Funding Agency**
To Improve the Program	To get the program on the right level and useful for the real world we live in.	To improve upcoming sessions.	To determine program strengths and weaknesses.	To get the program up and running and producing results quickly.
To Find Out What Happened in a Program- How It Had the Effects It Did	To find out what's being done in different work sites and share that with all the local unions.	To be able to say that training is effective because of how it was delivered.	To show the funding agency that the program has been carried out as stated in the grant proposal.	To document that the program met requirements of grant.
To Find Out About Short-term Effects	To show we are using what we learned to make our workplaces safer.	To show that students are learning something, and developing relevant skills.	To document the value of the program to union leadership.	To show that work places are increasing compliance with relevant OSHA regulations.
To Find Out About Long-term Effects	To show the company that the education programs are in their interest too.	To formally document that increased awareness created by the program is contributes to improving conditions at the workplace.	To determine if students have been catalyzed to take action, and are working together.	To show policy makers that the overall program approach leads to reduced injuries and illnesses.

The Need for Broad Participation: An Evaluation Advisory Team

This chapter presents a wide variety of options for the types of evaluations that can be conducted, for issues that can be addressed, and for methods that can be used. The point of view we take in this chapter is:

1. program evaluation will be strongest if decisions about the evaluation are consciously made by an evaluation team prepared to consider available options,
2. a well designed program evaluation will recognize and address issues posed by a variety of potentially competing interests head-on, and
3. an evaluation team should be formed and charged with helping those designing the evaluation with consciously deciding whose interests are to be served by the evaluation and which if any compromises will need to be made.

An evaluation advisory team can help define the evaluation's purpose, the questions it will address and its methods. When developing a team the program should consider how best to include the interests of:

- **workers** who are students in the program,
- the **funding agency,**
- the **sponsoring or participating organizations,**
- the education **program's administration,**
- the **program's instructors**
- **policy makers,** and
- others involved in the **health and safety movement.**

The most useful evaluations will make conscious decisions about whose interests are a priority and how a variety of interests can best be accommodated.

Three basic and interrelated questions must be answered in planning any program evaluation They are:

- **What kind** of evaluation will be conducted?
- **What questions** will the evaluation attempt to answer?
- **In whose interests** will the evaluation be conducted?

An advisory team can play central roles in the evaluation that will help ensure a genuine sense of ownership. The roles of the evaluation advisory team could include:

- **planning** the evaluation

- **guiding and advising** as the evaluation is developed and implemented, and
- **reviewing** evaluation materials and reports.

While balancing various interests is important, we must not lose sight of our ultimate interest in ensuring that the program evaluation and the questions it addresses are those that will contribute most to worker health and safety.

3. Two Approaches to Evaluation Methods: How Much and How Many or How, What and Why [1,2]

Quantitative methods address questions of how much or how many. They stress:

- **ratings and scales** - using numbered categories to measure attributes of the program or those affected by it,
- **objectivity** - making measurements, objective, valid and reliable in part by ensuring that the evaluator is not directly involved in the program,
- **statistical tests** - to make conclusions, and
- **generalizing** - from the sample of participants in the study to the program as a whole or to others in similar situations.

Qualitative methods, that is, more in-depth, open-ended approaches stress:

- **discovering** what is happening in a program, how and why,
- **the insiders view,**
- **the meaning of the program** to those who are part of it,
- **circumstances or context** in which the program operates and the participants live and work,
- **how peoples' views affect what they think and do**,
- **a subjective, yet unbiased point of view** taken by the evaluator who is close to the program,
- **gaining a deep, rich understanding** of people and their situation,
- **a focus on specific cases** that don't necessarily permit generalization to others,
- **a sympathetic understanding** of the program and those who take part in it.

Focusing on the Strength of Each Method: Using Both

A number of evaluators have tried to understand the strengths and weaknesses of both *quantitative* and *qualitative* approaches rather than focus on either one as the ideal. These evaluators have emphasized the need to use the methods that will:

- **be most useful** for answering the questions they have posed for the evaluation,
- **be most consistent with the principles and style** of the program and its sponsors,
- **provide results that are understandable** to those who will actually use them, and

- **best match the capabilities** of the program.

Evaluators that favor using both approaches believe that the strengths of one approach can be made to offset for the weaknesses of the other.

For example, some quantitative methods require a questionnaire or survey instrument. A good questionnaire may not exist for the evaluation questions being studied. Developing an instrument that meets "scientific standards" can be a major task. Using someone else's instrument may not address the specific question the evaluation needs to answer. More open-ended, in-depth qualitative methods can offset the limits of questionnaire writing.

Different Methods of Gathering Information

The different methods used to gather information and data for the two kinds of studies are shown below.

Quantitative	Qualitative
• **telephone, mail, or in-person surveys** using • **standardized questionnaires, observations, or tests** made up primarily of • **fixed-choice questions** where the possible responses are pre-decided and limited (for example, multiple choice or fill-in questions, or checklists)	• **direct observations** of the education programs or actions taken as a result of the program • **review of written documents** or records • **in-depth interviews** of individuals or groups using open ended questions that people can answer in their own words • **questionnaires** which supplement other data

Summary

While this chapter presents *qualitative* and *quantitative* methods separately, we believe the methods can and should be used together to strengthen evaluations. *Quantitative* methods will be best for finding out how extensive or general our findings are and *qualitative* methods will be best for looking in depth and finding out how significant our programs are for those that use them. We need to know both.[3]

4. Quantitative Methods

In general, quantitative methods seek to answer questions about the program by counting and assigning numbered values to answers to questions, and to observations or documents that are reviewed. Characteristics of quantitative methods are:

- **logic and objectivity**,
- **numbered measurements**,
- **surveys of groups**
- **questionnaires** with multiple choice, and fill in questions and checklists,
- **consistent and repeatable** procedures that are "reliable,"
- **accurate measurements** that are "valid,"
- **statistics** and a "hard science" orientation,
- **an outsider's view** for the evaluator, one that is detached from the program and at a safe distance from the data, and
- **generalizing** about groups.[4]

This section will address issues related to using quantitative methods to evaluate a program.

THE JARGON: Two qualities of survey questionnaires in quantitative studies are:

- **Reliability** - refers to how accurate the questions are in measuring what they are supposed to measure.

  For example, if the answers to questions changed depending on who asked them, how they were asked, or how tired students are when they answer them, the series of questions would not be considered reliable.

- **Validity** - for a questionnaire describes whether or not the questions are measuring what they are intending to measure.

  For example, if a questionnaire was used to find out how useful a class was for students, but instead was measuring how much students supported the organization that was sponsoring the program or how much they liked the instructor, the measure would not be considered valid.

Proof and Cause: Can We Find Them or Do We Need Different Questions?

One of the attractions to quantitative methods for impact or outcome evaluations is the hope that by being "objective" the evidence that is obtained can be considered *proof* or confirmation that the program *caused* certain changes. However, quantitative studies are generally required to make statements about *cause*.

Imagine the response to these requirements from someone responsible for a health and safety education program who already has too much to do and is trying to figure out how to squeeze in the time and resources necessary to conduct program evaluation on their own.

Here are the four requirements often used for determining that a program caused the effects that were measured.

1. The evaluation must begin with a statement about how the evidence is expected to turn out**,** a **hypothesis**,

No way! We're not getting into that.

2. The results must show a **statistical association** as predicted by the hypothesis.

We'll maybe calculate some averages, but that's it.

3. There is a way of making sure that the **changes happened after the program**,

O.K., we can handle that one.

4. **Other possible explanations** for the changes need to be systematically **ruled out**.

How are we supposed to do that?

The common "scientific" way to provide all these conditions is to use two or more groups to make comparisons in a study with what are called experimental or quasi-experimental designs. Some one-group designs are acceptable for looking at whether a program caused certain outcomes or impacts, however, the stronger of these designs typically requires that information be collected at least three times before and at least three times after the program.

Get real! If we wanted a study like that, and we could afford it, we'd hire a professional.

The world workers live and work in is complex. Just as with a chemical leak or spill, there is never a single cause for why things happen. This is just as true for changes that occur following an education program. There are often a number of causes for why changes occur. The task is sorting them out and ruling out other possible explanations for changes besides the program. So, while even the most "scientific" evaluation does not determine cause beyond doubt, you probably want to turn to an experienced evaluator who knows scientific evaluation

methods if you want to zero in on whether the program "caused" certain effects among a large group of participants.

Looking at Some Different Questions

If you choose not to get involved with these scientific standards for showing your program was *a significant cause* for improved worker health and safety, what other reasonable points of view can you take about using evaluation to show your program's effectiveness?

Perhaps we need to keep thinking about causes, but focus on some different questions - like these:

- **How can we reduce our uncertainty about the effects our program is having?** In other words, if we have doubts about whether or not a program is causing certain results, what can we do to reduce those doubts.
- **How can we better understand our program's contribution to those effects?**[5]

These less demanding questions require a different kind and level of evidence. Answers to these questions can help us see how the information we gather is or is not consistent with our thinking about how our programs should work.

Two evaluators who have written about how to conduct program evaluations, , Carol Taylor Fitz-Gibbon and Lynn Lyons Morris,[6] have said:

> *The critical characteristic of any one evaluation study is that it provide the best possible information that could have been collected under the circumstances, and that this information meet the credibility requirements of its evaluation audience. (p. 13-14)*

What they are suggesting is that we conduct the best possible evaluation for our purposes. Having said this, we also need to keep in mind those that will view our findings with a skeptical eye. Knowing and trying to limit the weaknesses of our evaluation can help us strengthen it and lead us to more reasonable conclusions.

Some Simpler Approaches for Answering Our Evaluation Questions

Seldom, do people on the front lines in carrying out health and safety education programs have the desire or ability to make such comparisons, especially on a "scientific" basis. More typically, program evaluations **rely on two types of information:**

1. **Information collected <u>after participants have taken part</u> in a program**, or
2. **<u>The difference between</u> information collected <u>before</u> a program <u>and</u> that collected <u>after</u> the program.**

In this last case workers are being compared to themselves rather than to another group.

For example, a questionnaire that asks workers about actions they have taken in dealing with chemical spills in the last six months might be given before a program and six months after the program. Comparisons would then be made between the two sets of answers.

These types of evaluation study designs may be most consistent with a program's information needs, its abilities and resources to conduct the evaluation, or simply with how a program wishes to conduct the evaluation. A brief description of several designs using these approaches will be presented with examples of how they were used. A brief description of some more complex designs will also be given.

Some Realistic Evaluation Designs

Each of the designs briefly presented in this section has been used to evaluate worker education programs.

Design 1: Comparing Information Collected From One-Group At Two Points in Time

This design provides for a comparison of a those participating in a program with themselves at some later point in time. Two types of comparisons can be made:

1. information gathered **immediately before** a program can be **compared with** that collected **immediately after** the program, or
2. information collected **immediately after** the program can be **compared with** that collected in a **follow-up survey three, six or twelve months after** the program.

This design allows evaluators to determine what factors have changed.

With the second design, there may be a problem with different rates of response for the survey at the program and the follow-up survey. Large differences in the rates of response make useful comparisons complicated. If evaluators are not cautious, biases created by differences between those who did and did not respond to the follow-up survey may affect conclusions that can be drawn from the study.

The following are two examples of NIEHS grant funded programs that have used the one-group design with two separate points of information collection.

In 1992 the Midwest Consortium for Hazardous Waste Worker Training reported on an evaluation study that asked participants a series of questions both before and after a program. The study found:

- increased student awareness and concerns about hazardous waste health and safety issues,
- changes in beliefs about the risks of hazardous materials, and
- heightened concerns about getting sick from exposure to these substances.

(Midwest Consortium report to NIEHS made by the University of Kentucky, June, 1992)

***

The United Auto Workers (UAW) hazardous waste worker education program compared workers' responses to a questionnaire completed at an education program with those gathered six months later. The comparison showed that there was workers improved their work practices including increasing their use of personal protective equipment and hazardous chemical information. (UAW report to NIEHS made by the University of Michigan, June, 1993)

A Special Case of Design 1: Comparing Information Collected at Initial Education Programs With That Collected at a Refresher Programs

The International Union of Operating Engineers (IUOE) adopted a unique method for making comparisons between two groups of hazardous waste worker trainees. They compared scores on questionnaires given to students about to take an initial 40-hour course with scores on the same questionnaires given to students about to take an 8-hour refresher course approximately one year earlier.

As an example, the International Union of Operating Engineers (IUOE) asked students to rate themselves on their ability to perform specific personal protective activities: decontamination and selection and donning of personal protective gear. Comparisons showed that those who were about to take the initial 40-hour course were less confident than those about to take the 8-hour refresher course. The evaluators suggested that this difference was consistent with students having learned and gained confidence in the initial 40-hour program that was retained over the 12 month period between initial and refresher programs.

The IUOE performed similar comparisons using scores on knowledge tests taken before the 8-hour refresher and the initial 40-hour courses. (International Union of Operating Engineers report submitted to NIEHS, June, 1994)

This approach provides a unique way of making comparisons without use of mail-back or telephone follow-up surveys. Conclusions from comparisons should be made with caution if the people in the initial and refresher program groups are substantially different.

Design 2: Relating Different Pieces of Information Collected From One-Group at One Point in Time

One of the most common designs used in evaluations is to gather information from a single group at one point in time and then examine different pieces of the that information to find possible relationships.

For example, a survey six months following a program might try to find out if students' actions to get needed work site health and safety changes are related to their ongoing use of the program's manual.

When relationships between two factors are found, most often this type of study does not permit evaluators to say that one factor led to the other. In other words, in the above example it couldn't be shown if use of the manual increased the likelihood that students would take action, or if taking action led students' to use the manual. Evaluators would only know the two were likely to occur together.

The United Auto Workers used information from a follow-up survey of students to examine the relationship between the program's impact on health and safety and workers' views of the company's health and safety "climate." Those who rated their work site health and safety climate more positively were more likely to report that:

- the training program affected work site changes in control measures, and

they followed recommended health and safety work practices. (United Auto Workers report to NIEHS by University of Michigan, June, 1994).

Design 3: Information From One-group Without Comparisons

There may be times when information from a single group is only gathered after a program. In this case, the evaluation may ask students directly about their perceptions of changes that occurred following the program. Because this group does not make comparisons it may not be considered as strong as the previous three designs for confirming that changes occurred. This design may be strengthened by using more than one group as a source for information. Every person interviewed in an evaluation has a unique point of view. Interpretations and memories of events that happened weeks or months ago can vary. When two groups of individuals with

different points of view provide similar information about what has happened it may help reduce uncertainty and add strength to the observations. The International Chemical Workers Union (ICWU) study of its NIEHS Hazardous Waste Worker Program provides an example of this approach.[7]

> The International Chemical Workers Union (ICWU) conducted a follow-up study of sites where workers had attended a hazardous waste education program. Their study had neither a comparison group nor pre-program information with which the follow-up survey results could be compared. Instead, workers were directly asked what had changed at there work sites in the 12 month period after their training.
>
> To help reduce doubts some might have about information solely gathered from union members about a union program a second point of view was obtained by asking the same question to a group of managers who had also attended ICWU education programs. Information from these managers provided those reading the evaluation report with an additional source of data they could use to draw their own conclusions.

Comparing Two Groups: Those Who Participated in a Program and Those Who Did Not

While for the most part using comparison groups in an evaluation is beyond the scope of this manual, we will very briefly present a couple of evaluation designs that may be possible. Making comparisons with different groups will create additional possibilities for looking at and understanding a program. However, using comparison groups will also make the evaluation much more complicated if the evaluation is to be done well and produce meaningful results. If you are considering conducting an evaluation that will compare groups you probably need the advice and help of a knowledgeable and experienced evaluator.

Design 4: The Delayed-Program Comparison Group

When a program is going to be provided to a number of different local unions, work sites or other groups and it will take a significant amount of time to reach them all, it may be possible to conduct the evaluation by comparing those who receive the program in an initial phase with those who will receive the program in a later phase. This design requires a great deal of up-front administrative work and that program planners have a good idea of who will attend the program over a period of time. It is often considered one of the fairest ways to conduct a comparison group evaluation as everyone receives the program more or less as they would have otherwise. This design also creates the possibility to assign groups to the different

phases of the program giving some control over keeping the groups as similar as possible. This similarity helps strengthen conclusions that can be drawn from the comparisons. The International Chemical Workers Union published an article describing an evaluation using this strategy in its cancer prevention program in the mid 1980's.[8]

Design 5: Comparisons With a Group That Did Not Participate in the Program

Sometimes it is possible to locate a group similar to those who participated in the program who either did not participate in the program or who participated in a different program. With this design there may be substantial differences other than the program between the participant and non-participant groups. If these differences could affect the impacts being studied, the differences need to be identified and steps need to be taken to answer why changes should be attributed to the program and not the other factors.

At one large facility where the United Auto Workers trained only a portion of the workforce comparisons were made between those trained and those not trained. Information was collected both before and after the training program. Before the program no differences were measured between those who received the training and those who had not. At the time of the follow-up evaluation, however, 47% of the trainees reported changing work practices, whereas only 18% of the non-trainees reported similar changes. (United Auto Workers report to NIEHS, June 1993)

A Final Note of Caution on Designs

The evaluation information we collect can serve a number of important purposes. Two of these are:

1. to reduce our uncertainty about the kinds of effects our programs are having, and
2. to gain a better understanding of how our programs might be contributing to the effects we're trying to achieve.[9]

Be careful about using quantitative methods to find "proof" that a program "caused" certain impacts or outcomes. Try to match your evaluation methods with a purpose you can achieve.

Surveys and Questionnaires[10]

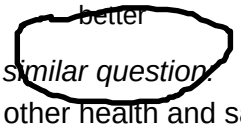
A prime method for gathering information in a quantitative study is to conduct a survey using a questionnaire. Typically, possible sets of answers for questions are determined ahead of time, and persons choose from among these sets of answers. Their responses are then

given numbered values. These values make it possible for evaluators to count or add responses of groups of people and to calculate statistical measures such as averages (or means). Those knowledgeable in statistics can calculate other measures of how much answers vary and the relationships between different factors measured in the questionnaires. The use of set answers for questions, that is, multiple choice, fill-in or checklists, provide different types of information than interviews that use open-ended questioning.

For example, consider the following response to a question on an end-of-the-program evaluation questionnaire and an open ended question in an interview to gather the same information. First, the fixed response question:

How would you rate this program compared to other health and safety education programs you have attended? Would you say it was: Much worse, somewhat worse, about the same, somewhat better or much better than other programs you have attended?

1. Much worse
2. Somewhat worse
3. About the Same
4. Somewhat better
5. Much better



Now consider a more open-ended interview pursuing a similar question:

How would you describe this program in comparison to other health and safety programs you have attended?

It was pretty good.

Can you explain what you mean when you say 'pretty good'? Did you like the whole program?

Well, the first part was pretty boring, but Anita, she did the afternoon session and she's got a great sense of humor. She's got a good style - more interesting.

Let me put it another way - we're trying to understand how useful you think this program will be for you?

Well, it's kind of a shame. I mean the stuff we covered in the morning is really important for the kind of work we do, but like I said, it was boring. Jim, I think that was his name, he seemed like a nice enough guy ... but now Anita's workshop, I've had that stuff ten times ... we do that at the plant twice a year, but nobody has ever presented it like she did. She's a good teacher. So I would say the afternoon was good, but as far as useful, something that I can really use, well ... I guess I would say it was just OK. You can always learn a little something. That's the attitude I try to take.

In an open-ended, person to person interview we can ask follow-up questions to get a clearer picture of what we're trying to find out. When we use written questions with specific choices for answers, getting a clear picture can be more difficult.

Getting the information you want from a questionnaire takes some hard work up front. Slight wording changes in a questionnaire or the order in which questions are asked can change the types of answers you get. The questions need to:

- **be clear and precise,**
- **cover enough angles of key concepts** so that they actually find out what you want to find out. If a topic is especially important try not to rely on just one question.
- **be carefully worded and asked** so that the answers don't depend on who is asking them and how they're being asked. To the extent possible everybody needs to understand the questions in close to the same way.

Whether the questionnaire or questions within it are your own, someone else's you have borrowed, or a combination or the two, the questionnaire should be reviewed with an eye for identifying potential problems and making improvements. The following list of review questions can be used for this purpose.

The Jargon:

- **Fixed Response Questions** are those with multiple choice, fill-in or checklist answers.
- **Open-ended Questions** are those where a person is free to use their own words to describe their response to a question

Questionnaire Content: Something Old, Something New, Something Borrowed ....

The selection of content for the questionnaire will be helped greatly by having a clear and specific purpose for the evaluation. This usually means having a small number of clearly stated evaluation questions the study will attempt to answer. Determining the major questions for the evaluation is covered in the sections of this chapter titled "Evaluating Implementation: What Happened in the Program" and "Evaluating Program Effects: Impacts and Outcomes."

For most worker health and safety education programs time, energy and resources for developing a questionnaire are very limited. Before you consider creating a new questionnaire think about using or adapting an existing questionnaire or set of questions. If you cannot find a set of questions that fits well with your evaluation you may need to consider creating a new questionnaire, overhauling an existing one, or making some compromises that enable you to make use of an existing questionnaire.

The National Clearinghouse for Worker Health and Safety Training is making available for review and use a collection of evaluation survey questionnaires that have been used by a number of worker health and safety education programs. **These questionnaires cover a broad variety of subjects including:**

- **ratings of education programs** by students,
- **use of written course materials** by students following the education program,
- **assessing students' knowledge and skills** related to hazardous materials and related health and safety subjects,
- **students' self ratings** of their own knowledge and skills,
- **handling of incidents** involving hazardous materials,
- **awareness, concerns and attitudes** about health and safety issues,
- **getting work site changes** in policies, programs, equipment or conditions,
- **work practices** related to health and safety,
- **workers educating their coworkers** following attendance at an education program,
- **ratings of train-the-trainers programs** by students'.

Some of theses questionnaires are designed for use immediately before and after programs, while others are for use at three, six and twelve month follow-up. Some are focused on individuals while others focused on work sites and changes in their programs.

Regardless of your source for a questionnaire, you should review it with an the intent of finding weaknesses and making improvements. Some of these may be general and others may be particular to the people involved in your study. The following checklist is provided as guide for this review.

A Checklist for Reviewing a Questions[11]

From the point of view of the full range of persons expected to answer the questionnaire:

- Are the questions easy to read?
- Are the questions brief?
- Is the meaning of each question clear?
- Are the questions and possible answers as specific as possible?
- Is the use of technical language at the right level?
- Do questions need to be translated or rewritten to make the questionnaire clearly understood by a different language, ethnic or cultural group?
- Does each item ask a single question?
- Does each of the possible choices for a response to a question provide for a single answer?
- Do questions avoid making bad assumptions?
- Do questions put all the possible answers on an equal footing?
- Do the questions avoid steering the person to certain answers?

THE JARGON:

- **A double-barreled question** is one that is really asking two questions at once with each question possibly having a different answer. Such an item may leave the person confused about which question to answer. For example:

What factors contributed most to your decision to use the quantitative and qualitative sections of this evaluation manual?

- **A loaded question** - one that uses loaded language that tries to bias the person to a specific set of answers.

What role do you think "so called" experts should play be in conducting evaluations of worker education programs?

- **A leading question** - one that encourages one set of answers over another.

  How did you like this evaluation manual?

  This wording may lead a person to think that they should have liked the manual. Some evaluators suggest using wording that lets the person answering know that you expect and accept that full range of possible responses.

  Some people have told us they didn't like the manual at all. Others have told us they liked it very much. How about you?

  Other evaluators claim that such wording doesn't make much difference. They recommend simply asking the question directly using a balanced approach.

Further Reviews and Pretests of the Questionnaire

Once questions addressing the major evaluation questions have been gathered, written and assembled into a questionnaire two additional steps need to be taken to test and refine it.

1. **Get feedback.** Seek the advice of coworkers who have experience in writing questions and questionnaires or who have an understanding of what you are trying to find out.
2. **Pretest.** Once you are comfortable with a draft of the questionnaire you should pretest it before using it for the actual evaluation. This is true whether you developed a new questionnaire or you modified or simply borrowed someone else's questionnaire. The pretest should be with a small number of people similar to those you will be asking to answer the questionnaire in the evaluation. To the extent possible, the way the questionnaire is pretested should match the way it will be used in the study.
3. **Conduct follow-up interviews.** As soon as possible following the pretest, contact those that completed the questionnaire and review it with them item by item. Ask them about their interpretation of the questions, any points of confusion and about their general reaction to the questionnaire. Get their suggestions for changes, including items they feel should be added or removed. Sometimes it is useful to conduct these interviews with a small group of individuals who have just completed the questionnaire.

   Use the pretest and follow-up interviews to check:

   - **the rate of response** to the questionnaire, ,
   - **the rate of completion** for the entire questionnaire and whether or not there are questions people regularly choose not to answer or say don't apply,
   - **length** - that the questionnaire takes a reasonable amount of time to complete,
   - **instructions** - that they are clear and are followed,
   - **understanding** - whether or not questions and responses are easily as you intended them,
   - **possible answers** - that options for answers to questions cover the range of answers people want to give and that any open ended questions will provide meaningful and useful answers.

If you modify a questionnaire following a pretest, if at all possible, pretest it again. Changes may solve one problem and create another.

Methods for Collecting Information

Quantitative information gathering frequently involves use of questionnaires. The following table presents some pluses and minuses regarding:

1. **written questionnaires (mail-back and at a program)**,
2. **personal interviews** using questionnaires, and
3. **phone interviews** using questionnaires.[12]

In-person, telephone and mail-back questionnaire methods may be combined to obtain a more cost effective and efficient survey. Often telephone surveys are used to follow-up on those that have not responded to mail-back surveys. Or those to be interviewed by phone or in-person may be sent an advance copy of a questionnaire so that they will have time to look over and think about the questions and gather needed information before the actual interview.

THE JARGON:

- **A Response rate.** This refers to the percentage of those given questionnaires that actually complete and return them. Response rates at an education program may be close to 100%, however, for mail-back questionnaires they can range from 10% to 90% with 50% or less being common.
- **Response bias.** Low response rates can lead to biases in the information collected. When those that complete and return a mail-back questionnaire are different in some important way from those who don't, the answers may lean in one direction and not be reflect the views of the whole group. This usually isn't a problem with a response rate of 80 to 90%. If the response rate is low, however, evaluators should be very careful about what they say about the information that may not represent the whole group.

Pluses and Minuses of Different Methods of Data Collection for Surveys[13,14]

	Mail-back Written Questionnaire	Personal Interviews	Telephone Interview
Response rate	Can range from 10 to 80%. 60% is acceptable in surveys of the general public. Higher rates are possible with highly interested groups.	80% is considered acceptable in surveys of the general public. Higher rates are possible with highly interested groups.	75% is considered acceptable in surveys of the general public. Higher rates are possible with highly interested groups.
Response bias	Low response rates can limit the ability to make generalizations and bias comparisons between information gathered in surveys done at different times.	Higher response rates will reduce the chances of response bias.	Higher response rates will reduce the chances of response bias.
Costs	Low cost especially if done at the education program. Costs for mail-back questionnaires may be $1 or more. Follow-up needed to increase response rates may increase costs.	High cost compared to mailback and telephone.	Moderate to higher cost. May range as high as $30 if conducted by a research firm. May be less if done locally with existing staff or volunteers.
Length	Over 12 pages may begin to lower response rates.	Lengthier interviews (over 1 hour) may be possible.	Over 1 hour may lower response rates.
Interviewer effects	None. There are some privacy advantages of mail-back questionnaires.	On the one hand, the presence of the interviewer may help motivate the person to respond. On the other hand, a person may try to give an interviewer what they think are desirable answers.	There are some of the same privacy advantages of mail-back questionnaires.
Ability to probe and clarify questions	None	For example, interviewer can read body language noting confusion, or respondent can ask the interviewer to clarify question.	For example, interviewer may be able to hear confusion, or respondent can ask the interviewer to clarify a question
Other		Visual displays may be used to help understanding	Telephone surveys can be mounted and completed quickly.

Comparing Information Collected at a Program With That Collected in a Follow-up Survey

Sometimes evaluations compare information collected from questionnaires filled out at a program and with information from the same types of information collected at a later date by phone or mail. A problem arises when there are large differences between the response rates for the two surveys. When nearly a 100% respond to the questionnaire at the program response bias isn't a problem. However, if the response rate for the follow-up questionnaire is much lower, say 50% or lower, the response bias may create big problems in making comparisons. If you have reason to be concerned about this issue you might want to seek the advice of someone who has experience in conducting surveys to help you deal with it (or see appendix).

For example, lets say you send a questionnaire to 100 people who attended your confined space entry program and only 50 mailed their questionnaire back. If 40 of those who mailed them back said the program was either good or excellent, that would be 40 out of 50, or 80% that said the program was good or excellent. But, what if those that liked the program were much more likely to respond than those that didn't? What if 40 of the 50 people that didn't respond would have told you that the program was poor or average with only 10 saying the program was good or excellent? That would mean that 50 out of 100 persons or 50% said the program was good or excellent. This gives a very different impression than the 80% positives among those that mailed their forms back.

Planning and Monitoring the Survey

Preparing Interviewers

Phone or in-person interviewers need to be acquainted with the program, its purpose and its methods. They also need to become familiar with the questionnaire, its purpose, and the range of possible responses they can expect. They need to know how to help put those being interviewed at ease, to identify and clarify confusion, and to record responses accurately. They need to be reminded to speak clearly. Interviewers should have a chance to practice interviewing in role-plays, and be involved in pretests of the survey.

Protecting the Rights of Those Who Participate: Confidentiality, Anonymity and Other Considerations

Ensuring **confidentiality** is one to protect the rights of those who participate in a program. Promises of confidentiality may also help persons speak more frankly. Promises of confidentiality should not be taken lightly. Confidentiality is protected when those gathering information

- have some method of connecting a responses with the person who provided them
- those who have the ability to make this connection is strictly limited
- lists or other means by which persons can be connected with their answers are tightly controlled, and stored in a secure location with access limited to evaluators, and
- reports of findings ensure that it is not possible to the responses of any one person.

Confidentiality is often improved by assigning each individual participant in the evaluation a unique number that is kept on a list with their name. The only personal identifier on the questionnaire is the person's unique number. This list is kept in a secure location separate from the questionnaires and data. Such a list can be very important in making follow-up contacts and for knowing who has and has not contributed to different parts of the evaluation (see the earlier discussion of response rates and response bias). Once the need to be able to connect the name with the questionnaire or data have passed, the list can be destroyed.

Anonymity exits when there is no way to connect a person with the responses they provided. No names or numbers are placed on the questionnaires. This is most easily accomplished with mail back questionnaires.

Confidential surveys provide advantages over anonymous surveys for follow-up. If questionnaires are anonymous there is no way of identifying who has not responded for purposes of follow-up. If information is to be collected at two points in time, names or numbers connected to responses enable evaluators to sort out possible response biases. The advantage of anonymity over confidentiality is that it provides more protection of privacy. There are trade-offs.

No person should feel coerced into responding to a survey or any particular question. This should be made clear to people when they are asked to participate. Evaluation surveys should include steps to help ensure that those participating understand:

- what information will be gathered
- how that information will be used
- who will and will not have access to their responses

- any possible negative outcomes from their participation in the survey

These procedures may help provide “informed consent” for those who participate.

Major institutions usually have boards whose task it is to review study procedures to ensure that they are ethical and that the rights of participants are protected. A review by trusted colleagues who are independent of your evaluation may help you identify and address issues of the rights of those from whom you plan to obtain information.

Managing the Survey

A well managed survey will ensure that information is properly collected and on schedule.

Announcing the survey. Pre-survey mailings may help alert people that a mail-back questionnaire or a phone interview will be conducted in the near future.

Phone Surveys. For telephone surveys, first mailing the questionnaire may give persons a chance to review the questionnaire and gather information or think about responses ahead of time. Timely follow-up is need for those who don’t respond. For phone surveys this means calling back, scheduling interview times, or leaving a phone number they can call (collect if necessary).

Mail Surveys. For mail surveys follow-up may mean sending reminder post cards, or additional copies of the questionnaire on a planned schedule.

For both types surveys, follow-up using the other method may prove helpful, that is, following up those who can’t be reached by phone with a mailing and those who can’t be reached by mail by phone. Refusals to participate should be respected.

Keeping and Storing Information. Once information is collected, keeping and storing records in an orderly fashion in a secure location are important. This is especially true if people responding to the questionnaire were promised confidentiality or anonymity.

Coding the Responses to the Survey and Maintaining the Quality of Information Collected

Once information is collected, it should be reviewed to ensure that answers have been properly read, interpreted, recorded and stored. If a computer data base is to be used, answers may need to be given a numbered value and entered into the computer. An additional review of data entered into the computer will be needed to find and correct errors. Questionnaires can be written so that some answers already have numbered values assigned to answers.

Statistics and the Analysis of information [15]

Statistics are estimates of characteristics of a group or population. Statistical methods are mathematical calculations used with quantitative data to:

- **describe** the data (for example an average),
- **compare** groups, or
- **examine associations** among specific items.

While statistics are useful to get a better picture of data from groups that would otherwise have to be described case by case, statistics has its own language that can seem alien and intimidating. While this manual will not attempt to teach statistics, a basic understanding of basic concepts may help evaluators consider options, know when to get help, and understand possible pitfalls.

THE JARGON:

- **Instruments or Measures** are the tools used to gather information or data. These often include questionnaires, checklists, interviews or observations, and tests.
- **Cases** are those persons, groups or organizations from whom data is gathered.
- **Units of Analysis** are similar to cases and describe the persons, groups or organizations for which the data is analyzed.

  *For example, program evaluation information may be gathered from a number of individuals from each of a number of different work sites. These individuals may be taken to represent only themselves as individual participants in the program or they may be taken to represent a group of participants at that work site. The information from the individuals could either be analyzed as **individuals who would be the unit of analysis,** or if the data from each work site are grouped or averaged, it could be analyzed as **work sites which would be the unit of analysis**.*
- **Variables** are measures of specific characteristics of the information gathered that may or may not vary among cases or over time.

  Age, gender, race, number of spills, ratings of usefulness of an education program, degree of concern for hazardous materials spills, number of coworkers trained, knowledge of chemical properties, could all be variables in an evaluation.
- **Sample and Population** - when the group of interest or the "target population" is defined, individuals from the group may be chosen as samples of the population.

Some Basic Descriptive Statistics

Descriptive Statistics are those that are used to characterize a set of data. The most common descriptive statistics are:

- **the range** - this is simply a description of the high and low values

  For example: "The size of hazardous materials spills reported in the six months immediately following training varied from 10 gallons to 10,000 gallons.

- **the mean** - is another term used to describe the average of the values of a group of data. It is a simple single value that describes a number of scores or values. It is easily calculated by adding all the scores and dividing by the number of scores. A mean doesn't always provide a good description of data. If there are a few values far different from the others they may present a distorted view of the data. These highly unusual values are called "outliers." Sometimes outliers should be remove and noted separately.

  If the size of 29 of 30 spills reported were between 10 and 50 gallons and the 30th spill was 10,000 gallons the mean size of a spills might be 360 gallons. This value does not describe the data well.

- **the median** - is the middle value or score. Sometimes it does a better job of describing a group of data. Half of the values should be equal to or greater than the median and half should be equal to or less than the median.

  The median of the 30 spills described above might be 32 gallons, giving a very different picture than the mean (average) of 360 gallons. A good description of these spills for the general reader might be: Thirty of the sites where workers attended the education program reported chemical spills in the first six months following the program. Of the 30 spills reported, 29 were between 10 and 50 gallons. The 30th spill was 10,000 gallons. The average for the 29 smaller spills was 30 gallons. Half of the spills were over 32 gallons.

- **variance and standard deviation** - describe how much variation there is in the data. A group of data with a large variance or standard deviation would be highly varied. Lower values for the variance or standard deviation would indicate that the values of the data are more closely grouped together.

Statistics for Making Comparisons

This section will not go into detail about using statistics to make comparisons. Persons interested in a more detailed discussion of statistics should look at book titled How to Analyze Data by Carol Taylor Fitz-Gibbon and Lynn Lyons Morris. This book is volume 8 of a 9 volume series on program evaluation and is available from Sage Publications in Newbury Park, CA or may be available at a local library. A very, very brief discussion may, however, be helpful.

A number of statistical tests can be used to compare groups or to see how certain items are associated with each other. Perhaps the most commonly used statistic for comparing groups is what is called the "**t-test**". For example, the t-test could be used to see if an average score for a measure of awareness of hazards for a group that attended an education program is different than the average score for a group that did not attend. When the comparison is made using the t-test there is a chance that a finding of either difference or no difference occurred by chance. When the probability that the finding was by chance is less than a certain percentage (often 5%), the difference may be referred to as a "**statistically significant**."

What They Mean, What They Don't Mean

Statistical tests, if used well, can help create a clearer picture of a large amount of data that otherwise would be a mystery. The use of statistical tests is, however, easily abused. The brief examples above show how statistics can create a distorted picture. Additionally, technical terms and "statistically significant" results do not make up for poorly designed studies, poorly worded questions or biased samples. **The first test of significance of an evaluation study should be whether the information from an evaluation is useful and helps improve our understanding of what happened in a program, the results it produced and how these contribute to the health and safety of workers.**

Three Kinds of Program Evaluation

Now that you better understand (we hope) some of the basics of qualitative and quantitative evaluation methods, we present three kinds of program evaluation for your consideration:

- **Using Goals and Objectives**
- **Evaluating implementation: What Happened in the Program**
- **Evaluating Program Effects: Impacts and Outcomes**

As you will note, there is overlap in these categories and the three sections of this chapter that address them. We leave it up to you to decide which approaches best suit your needs.

6. Goals and Objectives: One Basis for Evaluation of a Program

Goals and objectives are one way to define what a program intends to accomplish. Goals and objectives written about what effects a program intends to have can be used in to guide an evaluation of program impacts and outcomes. Those written about how these effects are to be brought about can be used to guide an evaluation the program's implementation. There are advantages and disadvantages to building an evaluation around goals and objectives. Thus, while we present the use of goals and objectives for the purpose of evaluation, other alternative approaches are presented in each of the other two sections about kinds of evaluation.

Pluses and minuses of using goals and objectives as the basis for program evaluation

Pluses

Using goals and objectives has some obvious advantages.

1. **Coordination and integration** - Potentially all aspects of the program are coordinated and integrated at the beginning of the program. Evaluation is not an afterthought.
2. **Satisfying the funding agency** - If an agency decided to fund the program on the basis of a proposal that included a highly detailed and well thought out program plan, an evaluation following this plan should largely satisfy the funding agency.
3. **Conclusions and report writing** - The goals and objectives can serve as an excellent basis for developing conclusions and for writing a report.

Minuses

Using goals and objectives as the basis for the evaluation also has some potential disadvantages.

1. **Assumes a program is well developed and not still evolving** - Basing an evaluation of program effects or how they were reached on goals and objectives may lead to defining the scope of an evaluation before a program has had a chance to become fully formed and stable. Goals and objectives of a new or forming program may need to be changed because of the experiences of the program. If the evaluation is already set, program planners may be reluctant to make changes that will affect the evaluation. In such a case the program can end up serving the evaluation rather than workers.
2. **Assumes the existence of a detailed program plan** - This approach assumes that all programs have the time and resources to undergo a detailed program planning process. This is not always the case. Some very successful programs function on an understanding of needs. These programs use plans largely borrowed from their own or others' past experience. Often these plans don't include highly refined and measurable goals and objectives.
3. **Sharing power** - Some programs are based on sharing power among those delivering a program and those participating in it. These programs are designed so that participants bring their own definitions of problems, share them with others and help shape and reshape the goals and objectives of the program and what students want to achieve both during and after the program.
4. **Missing the unexpected** - An exclusive focus on goals and objectives may steer an evaluation away from looking at program activities or results that weren't planned. Sometimes the most meaningful experiences in a program or the most important results are those that program planners didn't anticipate.

Setting Goals and Objectives[1]

One method for developing goals and objectives or examining existing ones for their usefulness for an evaluation is described below. The first four steps present a process for developing objectives for the *effects* a program is intended to create. A fifth step is then added that can be used to develop objectives for *how the program is to achieve these effects*.[16]

Setting Goals and Objectives for Targeted Impacts and Outcomes

Step 1. Develop Goals - Goals focused on program impacts and outcomes are usually broad statements about what a program is trying to achieve or a problem the program is trying to solve. Typically they are limited to one or two realistic statements. Write goals so that they clearly state:

- the health and safety conditions to be changed, and
- the people for whom the changes should occur.

For example, a health and safety education program goal might be: To reduce the number of confined space injuries among Ohio grain mill workers.

Be sure goals fit with the purpose and mission of the organization. Unlike objectives, however, goals do not need to be measurable or achievable during the length of a single program. They express long-range aims.[17]

Step 2. Brainstorm Indicators - Have the evaluation advisory team or others involved in planning the evaluation brainstorm a list of specific, observable or measurable factors (such as knowledge, skills, attitudes, actions, the results of actions or other factors) which you believe are logically connected to achieving the stated goal. These factors should be considered *good indicators* of important impacts or outcomes and progress towards the goal. If you have an idea or theory of how the program should work, this is where it comes in.

---

[1] **The process for setting goals and objectives is adapted in large part from: Four steps to better objectives. Wisconsin Department of Human Services, 1981.**

Think about indicators over time:

- short-term - during or immediately following the program,
- intermediate term - initial actions workers might take, and
- long-term - results of worker actions that may take some time to bear fruit and that are closer to indicating significant progress to the goal of reduced injuries and illness

The following is a *partial list* of indicators that might come up in a brainstorming session for the example of the goal given above.

Short-term

- *workers understand simple and chemical asphyxiation, oxygen deficiency, ... and so on*
- *work site teams can demonstrate the ability to review critical aspects of a confined space program*
- *attitudes change about the dangers of working in confined spaces*

Intermediate term

- *programs are put in place at each work site to educate the entire workforce about confined spaces*
- *labor-management committees are created to review current confined space programs and policies*

Long-term

- *all work site confined spaces are properly labeled with warning signs*
- *ensuring changes are made in equipment, policies and programs to eliminate confined spaces or to reduce the need to enter them*
- *trained on on-site confined space rescue teams are present at each work site*
- *all equipment necessary for confined space rescue is available*

Step 3. Select indicators - Group and combine those indicators that are very similar or have a high degree of overlap. Then, pick a few of the best indicators that could be looked at in the evaluation. Think about keeping the evaluation manageable. Base your selection of indicators on:

- importance for the various parties who have an interest in the program,
- easy to measure,
- likelihood that the indicator will change in the time frame of the evaluation.

Step 4. Writing outcome objectives - Translate the indicators into measurable objectives by having at least these three elements in each objective:

- **what** - a description of the indicator (chosen in step 3)
- **who** - the indicator is in reference to,

- **when** - the objective will be accomplished.

Another element may be added if you believe you have the ability to know:

- **how much** - change or achievement is expected.

Another additional element may be added once you have determined how achievement of the objective is to be measured.

- **how to measure** - achievement of the indicator

The following example includes all the elements except how much change is expected.

> *By 1/15/97 all of the 125 grain mill workers at 35 sites who have attended two-day confined space entry programs will have demonstrated that they could identify confined spaces, their potential hazards, the elements of an effective confined space entry program, ... as measured by observation of students successfully completing all educational exercises and simulations.*

This next example includes all the elements except how the result is to be achieved.

> *By 9/1/97 routine confined space entries will be reduced by 90% as measured by comparing pre-program questionnaire responses with 12 month follow-up phone interviews with participants at each site.*

Estimating the degree to which something should be accomplished in order to indicate success can be difficult. You may be able to estimate a targeted amount of change based on experience or what other programs have done. Often, however, an estimate must be based on a best guess, or what those in the program feel would mean the program was a success. If achievement of an objectives is to be a primary measure of the program's success, **be realistic**, try not to overestimate.

Setting Goals And Objectives For <u>How</u> Programs Will Achieve Their Effects

Process objectives state what needs to be done to accomplish the outcome objectives (see above).

Step 5. Writing Process Objectives - State process objectives by specifying *how* outcome objectives are to be achieved. Include in the objectives:

a) **who** - will participate in the activity (specify how many will take part)

b) **what** - describe what they will do or receive in the program,

c) **when** - will the activity take place or be accomplished.

By 6/1/97 15 of the 35 grain mill sites will have reported formation of labor-management teams to review their confined space entry program as measured by six-month follow-up phone interviews with program participants at each site.

Goals and Objectives for Measuring Success

Once set, the degree to which goals and objectives are attained becomes an important measure of program success. Evaluators interpret data and information they collect according to the goals and objectives, they reach conclusions by comparing accomplishments to what was stated would be achieved, and they create program reports using the objectives as an outline.

When goals and objectives are well written, this approach helps ensure:

1. program evaluation is focused on the aspects of the program that are most important to the interested parties
2. for grant funded programs that specify goals and objectives in their proposal, that findings will satisfy, at least in part, the needs of the funding agency
3. that there is consensus among the interested parties

Using Flexible Objectives as the Basis of Outcome Evaluations[18,19]

Two NIEHS funded worker education programs have developed flexible program objectives by using work sheets called risk charts during the training. The risk charts are a tool to help workers identify where there are and are not problems in specific areas of a company's health and safety programs, policies, procedures and equipment. Once students have reviewed and graded the company's program, they use the charts to develop priorities and plans for getting work site changes. These same categories are used in follow-up evaluations of program impact.

For example, in one follow-up impact evaluation several months after the program, students were asked whether they attempted to get the company to make changes in eleven targeted areas. If they reported that changes were needed and attempted, then they were asked whether the changes were obtained, or what their status was.

Goals and Objectives: Whether you use them or not

Some program planners and evaluators use the strengths of the goals and objectives approach as the basis of the evaluation. Even with the potential weaknesses many evaluators find that goals and objectives can provide a useful footing for thinking about how a program is operating, what effects are to be measured, and how we are to interpret results. Often, however, program and evaluation planners do not wish to go through the process of developing highly specific goals and objectives for the purpose of the evaluation. The implementation and impacts evaluation approaches and methods described in the next sections of this chapter can be used *with or without* goals and objectives as their basis.

7. Evaluating Implementation: What Happened in the Program[20,21,22]

Often we have a vision of how we want a program to work. Making that vision a reality is something different. Most programs don't go exactly the way we wanted. Sometimes what we plan doesn't work. The following real world example of an implementation evaluation of a cancer prevention program carried out by a major industrial union demonstrates the point. While this case may be different from your programs, the ideas that are applied to this case throughout this section should be able to applied generally to a variety of program types and subjects. This example will be referred to throughout this section.

A Cancer Prevention Program Case Study

The Program and Major Findings of Program Effects

In the early 1980's a major union cancer prevention education program was undertaken among rubber workers, a group known to have high rates of cancer. An evaluation of the effects of the program found little if any changes when thousands of workers were surveyed at group of 13 sites that had received a special Train-the-Trainers program were compared with 10 sites that had not. The education program was designed to stress three ideas:

1. *tailoring* education programs to each work site's specific needs,
2. *involving* workers in planning and implementing the programs, and
3. *partnerships* that relied on social supports among families and coworkers needed to help make changes.

The targeted outcomes were: changes in workers' knowledge, attitudes, beliefs and behaviors related to cancer prevention. Why were no changes found? Were the ideas behind the program, tailoring, involving, and partnerships, the wrong way to go, or was something else at work?

Implementation Evaluation Studies

To help answer these questions two types of studies were undertaken. An in-depth case study was conducted at one plant after the completion of the project, a plant that program personnel believed would have been most likely to succeed. In addition, at all the study sites information was also collected about program activities and educational events. The case study consisted of:

- three site visits,
- interviews with persons who played key roles in the program,
- observations of program activities, and
- examination of a large number of program documents.

Program monitoring data were collected monthly from all study sites and included information on:

- program planning,
- advice or help needed,
- program activities,
- problems and barriers encountered, and
- solutions applied.

Findings of the Implementation Study

The case study and analysis of the program monitoring data showed that there were several unanticipated barriers to effectively carrying out the program, problems that could have caused the lack of program effects. They were:

1. During the project OSHA implemented its Hazard Communication Standard. Most of the time, resources and energy available to address health and safety issues went into putting a hazard communication program in place, rather than towards developing an effective cancer prevention program.
2. The project lifespan paralleled a major decline in the industry targeted by the program. This included plant closings, and company buyouts and mergers. People's attention was directed at more immediate jobs issues.
3. Very few of the desired program activities were actually conducted. When planned activities were carried out, participation rates were low.

The Meaning of What Was Found

In other words, any assumptions that the program was carried out as planned and under normal conditions would be false. A conclusion that the program wasn't successful because of the program design would be unfair. This example clearly illustrates that we need to know how and why a program has succeeded or failed to fully understand the meanings of an evaluation of the program's effects.

This case study will be referred to throughout this section to help explain ways in which you can think about doing an implementation evaluation.

This case study is taken from: Allan Steckler, "The Use of Qualitative Evaluation Methods to Test Internal Validity," Evaluation and the Health Professions, Vol. 12, No. 2, June 1989, pp. 115-153.

Two Types of Implementation Evaluation

In general, evaluations that look at how a program works are divided into:

1. those that aim to help improve the program while it is still developing, and
2. those to find out how the program achieved what it did (often when it is over).

Evaluations Aimed at Improving the Program

The central purpose of this type of evaluation is to identify where program adjustments need to be made **to improve a program or keep it on track**[23]. This type of evaluation might also serve to:

- prepare for a new phase of the program, or keep a program description up-to-date.

Activities involved in evaluations to improve a program may include:

- pre-testing or piloting programs, their materials, and methods, or ongoing monitoring of program activities.[24]

Evaluations Aimed at Finding Out What Happened in a Program

While similar in many respects to evaluation used to improve a program, this type of evaluation commonly has a different time perspective and purpose. In general, these evaluations look back at a program with the purpose of better understanding **why a program had the effects that it did**. They answer how, when, under what conditions and with which groups a program approach will work.[25]

THE JARGON: Evaluations of program implementation include:

- **Formative Evaluations -** collecting and using information to shape or modify a program as it is getting started or ongoing.
- **Process Evaluations -** collecting and using information at the end of a program or at the end of some major phase of the program to find out what actually happened in a program - how it had the effects it did.

Program Framework, Delivery and Context: The Basis for Implementation Evaluation[26]

Both evaluations conducted to improve a program and those undertaken to document what happened in the program are commonly designed around three basic aspects of the program, its:

1. **The framework** - Is the design of the education program likely to lead to the desired immediate changes for workers, and are these changes likely to lead longer term changes and to reduced injuries and illness?

   For example:
   - ***if you believed*** *improving worker health and safety is best gained by creating changes in work site health and safety policies, programs and controls, and* ***if you believed*** *that such changes can be promoted by social actions taken by workers who posses knowledge, skills and the confidence to use them,* ***then*** *you might have designed your education program using small group, participatory problem-solving exercises led by workers that sought to provide needed information and develop and practice problem solving and social action skills.*

   Taking another approach:
   - ***if you believed*** *that improving the health and safety of chemical emergency responders is best addressed by improving how emergency response actions are carried out, and* ***if you believed*** *that such response takes highly developed individual and group skills carried out in a well designed emergency response plan,* ***then*** *you might have designed your education program using interactive lectures and hands-on, response action practice sessions led by highly skilled emergency response professionals.*

   In each case the ultimate goal is improved worker health and safety. In the first case the immediate objective is developing of problem solving and social action skills. In the later case the immediate objective is improved skills related to chemical emergency response. In each case you, the program planners, have chosen educational methods which you believe best suit the immediate and long term goals and objectives.

2. **Program delivery -** Is the program being delivered according to the principles and design planned into it? Think of how the cancer prevention program education program in the case study was not carried out as planned. This is a common problem.

3. **Program context** - Would this same program carried out under different circumstances be more or less effective? Again, think of the cancer prevention program case study and how the context affected the results.

When a program fails to deliver the expected results we need to know in what ways its framework was responsible. To know this we must understand to what extent the program was delivered as planned. We must also know in what ways factors surrounding the program either supported or hindered it. Then we can have a clear picture.

Understanding a Program's Characteristics

An implementation evaluation, whether looking at improving a program (*a formative evaluation*), or looking at what happened in the program (*a process evaluation*), can become a method for better linking the means the program uses with its desired ends. A program's **framework, delivery** and **context** are related to two equally important key types of program characteristics.[27]

1. Key Characteristics of the Program's *Activities*

These would include characteristics of the program as planned and as delivered. They might include:

- its **principles**
- its **methods and procedures**,
- the type and level of **student participation**,
- its **design**
- its **administration**,
- **interactions** among people and organizations involved in the program.

The cancer prevention program case study was built on the framework of tailoring programs to each work site, involving workers in planning and implementing the programs and partnerships and social supports among families and coworkers. It used a train-the-trainers design.

Thinking in terms of characteristics of program activities, what might you want to know about how that framework served the program?

2. Key Characteristics of the Program's *Context*

These would include the circumstances surrounding the program as well as the characteristics of those involved in the program. These might include:

- the program's **setting**,
- **characteristics of program staff or the students** who participate,
- the **political environment** within which
- the **size** of the program or its classes,
- the program's **materials, equipment, or other resources**,
- **criteria or eligibility** for student

the program operates, participation,

- workers' **employment situation**,

For example, do instructors and students have a good enough understanding of each other so that they can communicate and work together effectively?

What is the status of labor-management relations related to health and safety, or how do economic trends affect local union priorities and program and policies targeted for change?

Which characteristics are most important for your program? If you could only focus on one or two in an evaluation which would you want to look at first?

Evaluations asking about these characteristics are trying to find out how these factors combine to make a program more or less successful. They are looking at the overall program to see whether there are problems with 1) the framework, 2) with how the program was carried, or 3) whether other approaches are need to address barriers created by the context of the program. Program success depends on answering these questions to reduce uncertainties, to discover useful solutions to program problems and to properly interpret impact or outcome evaluation results.

Levels of Program Action - Targets for Program Evaluation

The short-term objective of a health and safety education program may be to help workers change their level of knowledge, skills, actions. Other short-term objectives might also include helping workers explore and change their beliefs, attitudes, or norms. Often program evaluations either stop at the individual, short-term level, or they move immediately to focus on longer term program effects. Of equal importance for understanding how a program works are the processes that participants and those they affect go through to carry out the program.

Imagine you were involved in an implementation evaluation of the cancer prevention program in the case study described earlier in this section. Recall that it used a train-the-trainers model. You could choose to look at how and why the program worked on three levels.

1. ***the initial train-the-trainers program***
2. ***worker-trainers conducting programs in the field***
3. ***organizational supports and issues***

When you consider the evaluation questions you want to answer, think of the various levels on which your program is supposed to act. These levels are important for both *formative evaluations* conducted to help improve an ongoing program, and for *process evaluations* conducted to better understand program effects.

Zeroing in on the Levels of Program Action

Be realistic. You may want to focus on those one or two levels at a time - the ones you most need to know about now.

The **Levels of the Program Action Work Box** has been filled in below as it might have been for the cancer prevention program case study. Another blank form for your use in planning your own evaluation follows this section.

◊ Levels of the Program Work Box

Briefly describe what levels of the program you wish to investigate in this evaluation.

⇒ **The initial education program:**

What were the perceptions of both instructors and worker-trainers about how well the train-the-trainers program helped prepare worker-trainers to conduct the program in the field?

⇒ **Workers' activities following the program:**

What specific actions did worker-trainers take to organize and conduct the programs at their work sites. Why did they or did they not engage in actions recommended by program planners?

⇒ **Other organizational levels (for example, the health and safety committee):**

How and why management did or did not support the program? Similar questions could be asked about the local union leadership? How and why were labor-management relations related to health and safety affected?

Determining Scope of the Implementation Evaluation

The scope of the evaluation needs to be manageable and realistic. An evaluation to answer all possible program improvement questions will be too large an undertaking. The list of evaluation questions in the **Selecting Implementation Evaluation Study Questions Work Box** (see the next page) can be used to help narrow the range of issues you consider. Copies of the work box can be distributed to all those contributing to the planning of the evaluation for ranking the questions according to their importance

Use the rankings to help:

- **sort through common and divergent interests** among those helping to plan the evaluation
- **build a consensus** around which questions need to be answered in the current evaluation.

Scores should be tabulated from all those completing the forms and reported. Those question scoring highest should be discussed further for consideration for inclusion in the evaluation. The aim should be to address a reasonable number of evaluation questions. One of implementation questions may be enough for the resources available. Perhaps the program can study more.

Think about: The Time, Energy And Resources Available for the Evaluation

As you proceed through the evaluation planning process, continually ask yourselves whether the time, energy and resources to conduct the proposed evaluation are available and whether answering the questions will be worth the effort that will be required. Studying a few issues well will be much more useful than studying a large number poorly. You don't want to overload the evaluation so that it becomes a burden to the program than a help.

Selecting Implementation Evaluation Study Questions

In using the **Selecting Implementation Evaluation Study Questions Work Box** think about:

a) **what you already know** and don't really need to spend time finding out (a ranking of 1 on the scale),

b) **what you need to know now** to improve the program, (a ranking of 5 to 3) - note: make any evaluation requirements of the funding agency clear up front), and

c) what you may want to know sometime in the future (a 3 or 2).

◊ Selecting Implementation Evaluation Study Questions Work Box

What priority would you give the following evaluation questions for your program? [28]	**Priority: 5=Highest; 1=Lowest**
1. **Strengths and Weaknesses -** What are the program's strengths and weaknesses? How can they be built or improved upon?	5 4 3 2 1
2. **Program Delivery -** In what ways is the program being delivered differently than was planned?	5 4 3 2 1
3. **Students' Views -** What are students' reactions to the program? Is the program on target for their needs? Is it at the right level?	5 4 3 2 1
4. **Language -** Is it too simple, or too difficult or technical? Does the language of the program match student's primary language?	5 4 3 2 1
5. **Views of Instructors and Other Staff -** What are the reactions of instructors and other staff to the program? What additional information, skills or resources do they need?	5 4 3 2 1
6. **Program Context -** What about the program's setting, curriculum materials, and methods? In what ways are they or are they not working?	5 4 3 2 1
7. **Target Audience -** Has the program actually reached those it is supposed to? Does program's effectiveness vary with the background of participants?	5 4 3 2 1
8. **Outside Influences -** What about the circumstances within which the program is operating? What forces outside the program are affecting how the program works - at the union, the work site or other?	5 4 3 2 1
9. **Program Objectives -** How does the program need to be changed to meet its objectives? How do the program's objectives need to be changed to be more on target?	5 4 3 2 1
10. **Program Balance** - In what ways does the program's balance between structure and flexibility need to be shifted?	5 4 3 2 1
11. **Program Impacts** - What kinds of affects might the program have, realistically? How does the program need to be changed to achieve these effects? How do expectations need to be changed?	5 4 3 2 1
12. Outcome Evaluations - What issues should be addressed in follow-up evaluations? What evaluation methods and systems will work best in examining program delivery and effects?	5 4 3 2 1
13. **Other:**	5 4 3 2 1

Refining Implementation Evaluation Questions

Following a preliminary selection of implementation evaluation questions the evaluation planning group should either rewrite the question or add detailed notes to make it more specific to your program. One general evaluation question may generate a number of specific question as the general question is applied to your program. Thinking about the cancer prevention program case study, lets say your evaluation group zeroed in on the Question 2 from the Selecting Implementation Study Questions Work Box about program delivery. You might fill out the **Refining Implementation Question Work Box** in the following way (another blank form for your use follows this section).

◊ Refining Implementation Question Work Box

Question from the Selecting Implementation Study Questions Work Box:

In what ways is the program being delivered differently than was planned?

Rewrite or add detailed notes to make the selected evaluation questions specific to your program:

⇒ *Who have worker-trainers tried to bring into the process of planning work site education programs? How have they done this? What's worked and what hasn't?*

⇒ *What specific planning steps have been used? To what extent have worker-trainers followed the step-by-step planning guide presented at the original train-the trainers education program? To what extent have they found the planning guide useful? What are the strengths and weaknesses of the planning guide?*

⇒ *When work-site education programs were presented, to what extent did worker-trainers use the methods, materials and curriculum in their program presentation kit? and so on ....*

Again, think about time, energy and resources it will take to answer these more specific questions. Keep asking yourselves whether the scope of your evaluation focus is reasonable.

Once the evaluation team has agreed upon the evaluation questions to be studied, the evaluation plans can begin to focus on methods and sources for data collection.

Choosing Data Collection Sources and Methods For the Evaluation[29]

Implementation evaluations may use a variety of methods and sources to gather information. Program descriptions, as well as conclusions or recommendations that are part of an evaluation will be **strongest** if the information they are derived from comes from a **variety of methods and sources**.

Sources of Information

Information may be collected from those either directly or indirectly involved in the program. For example:

- instructors,
- other staff,
- advisory groups or boards.
- shop stewards,
- coworkers of those who are part of the education program,
- other local union leaders,
- company health and safety personnel, or
- supervisors.

Consider first those sources of information that are already being gathered or that will take the least amount of time, energy and resources. Identify gaps. Think about the need for a variety of perspectives. Again, as you consider what other sources might be added think about the time energy and resources you will need for each additional source of information.

Methods of Information Collection

There are various options available for data collection for implementation evaluations. [For a more in-depth discussion of sources and methods of information collection refer to sections of this chapter on quantitative and qualitative methods.] Options for methods to collect information may include the following:

Direct Observations

Evaluators take detailed notes or complete checklists while viewing program activities.

For example, evaluator observations train-the-trainers classes, or planning meetings.

Written Documents

Evaluators take detailed notes or complete checklists while reviewing program documents.

• *grant proposals*	• *work sheets or flip charts,*	• *phone logs, or*

• *manuals*	• *registration forms,*	• *minutes of meetings;*

Questionnaire Surveys

Evaluators or other staff or conduct telephone, mailback or in-person surveys using primarily multiple choice, yes/no and checklist types of questions.

• *immediately <u>before</u> the program*	• *immediately <u>after</u> the program,*	• *at follow-up (3, 6 or 12 months)*

In-depth Interviews

Evaluators conduct interviews with individuals or groups using primarily open-ended questions,

• *in person individual interviews*	• *interviews with groups of 6 to 8 (focus groups)*	• *telephone interviews*

Monitoring System

Evaluators administer an established and regular monitoring system.

A checklist used as part of an instructor peer evaluation program for those conducting train-the-trainers programs that documents that key topics were covered, that key educational objectives were met using recommended methods.

Sources and Methods of Information Collection

The following is a **Sources and Methods of Information Collection Work Box** to help you decide which methods of information collection to use, possible sources of information, and when where and by whom the information would be collected. A sample has been completed for the cancer prevention program case study. Another blank form for your use follows this section.

◊ **Sources and Methods of Information Collection Work Box**		
Evaluation Question: *Who have worker-trainers tried to bring into the process of planning work site education programs? How have they done this? What's worked and what hasn't?*		
Possible Methods	**Details on sources and methods of information collection.**	**When, where and by whom will the information be collected?**
Direct observations of program activities including site visits	*One visit each to two work sites.*	*Two months after train-the-trainers program. Site must be close so visits can be done in day trips. Program Director or Lead Instructor.*
Questionnaire surveys	*Written questionnaire (borrowed and modified) given to all worker-trainers.*	*Immediately before the train-the-trainer program. At the training site. Brief two veteran worker-trainers and have them administer the questionnaire. (no staff in room)*
In-depth interviews of individuals or groups	*Two interviews session of groups of 6 to 8 worker-trainers*	*At annual trainers meetings. Conducted by Lead Instructor*
Reviews of existing documents	*Review of completed education session worksheets and flip-charts.*	*Request from at least two work sites each where staff feels worker-trainer led programs are more, and less likely to be successful.*
An established and regular monitoring system	*Report forms with sections on planning, activities, advice sought, problems encountered and solutions tried.*	*Monthly - completed and mailed in by worker-trainers with follow-up by Program Director.*

Analysis of and Use of Information[30]

Information collected for the evaluation should be analyzed according to the methods used. Scores, averages or other statistical measures should be computed from quantitative data sources. Major ideas or themes should be identified and organized from qualitative data (see previous sections). Drawings, graphs or charts can be used if they help clarify results and are an effective way of communicating for those receiving the report. The major findings and the supporting evidence should be organized and developed into an oral and/or a written presentation.

In planning for the analysis keep in mind:

- who needs to be informed about the evaluation process, its purposes and activities, and
- how will they be informed, kept up to date and involved in the process.

The Meaning of Information Collected[31]

Meaning is what one makes of the information that is collected to answer the evaluation questions. This phase is used to examine:

1. whether changes in the program plan are useful or harmful to the program, or
2. how the way the program was carried out might have influenced the effects that have been measured.

Evaluations Aimed at Improving the Program

Meaning. In evaluations to improve a program, evaluators must first consider the possible consequences of what is found. What does the information collected in the evaluation mean?

For example, if the program is being delivered in some way other than that planned, how do those changes either strengthen or weaken the program? How do the changes alter what the program is likely to accomplish?

If students enrolling in the program are not hose the program was intended to reach, is the program achieving its aims?

Suitable responses. Examination of questions like these should lead to discussions among the evaluation team, administrators, instructors and other staff about the possible suitable responses.

For example, should the program plan be modified to incorporate new ways instructors have found to deliver the program or should program delivery be brought back in line with the original design? or

Should the way students are enrolled be changed to better target those the program was originally designed for or can the program be viewed in a new way so that it includes all those who are expressing interest? and so on.

Deriving Meaning and Finding Suitable Responses

Meaning. Most often evaluations to find out what happened are trying to make connections with the information that is gathered in an evaluation of program effects. Thus, the process of determining what the information means is carried out using the findings of both types of evaluation at the same time.

For example, some students in a cancer prevention train-the-trainers program may have reported that they followed all the recommended procedures for organizing and delivering formal education programs back at the union hall but that very few members attended the programs. At other locals worker-trainers may have reported that the programs were highly attended and deemed successful. These findings might lead evaluators to analyze information from an impact evaluation survey separately for those sites that successfully carried out programs in the field and those that did not. If positive findings about the program's impact on changes in cancer prevention knowledge, awareness and actions were higher among rank and file at those locals that successfully delivered the program then a evaluators might conclude that the program could work if properly implemented. If no difference in effects was found, then one might wonder if the program was really an effective one.

A report that the program worked well where implemented well, but that some sites had trouble implementing the program would be much stronger than if a report lumped all the data on effects together and concluded that the program was modestly successful, the positive results averaged in with the negative, for the entire group of sites.

Suitable responses. Examination of this information should lead to discussions among the evaluation team, administrators, instructors and other staff about the possible implications for future programs or continuation of the current program.

For the above example, if findings of program effects matched findings about program delivery in the field, planners might choose to collect additional information on the differences and similarities of sites where the program did and did not come off well. Program planners might also choose to plan a future program keeping much of the same program content and materials but offering worker-trainers a variety of other options for getting information to workers on the shop floor.

Implementation Evaluation Reports

Implementation evaluation reports are only helpful if they are used. The ultimate success of an evaluation aimed at program improvements depends not just on the evaluators, but also on those who will implement program changes or be affected by them. Thus, an important aim of the evaluation is consensus among these various interests. Representatives for an evaluation team may include program advisors, administrators, staff, instructors, students and other key persons or representatives. To serve these various interests reports should be made in three steps:

- **The initial report** should be created by the evaluators and -
- **shared and discussed** with the evaluation team which works towards agreement on modifications to the -
- **final report** to be made by the evaluators.

Different users may have different interests in terms of an evaluation report. Reports for formative evaluation range from verbal presentations followed by a brief written report of conclusions to more detailed written reports that includes both conclusions and supporting documentation. The report for a process evaluation may need to be written to correspond with an evaluation report on program effects. The report may need to be included as part of an accounting process for the funding agency. The format for the report and who it is being prepared for will influence the types and sources of data, how the data is to be collected and how it is to be reported.

Think about:

- **who** will use the reports,
- **when** they will need them,
- **how** the reports will be used, and
- **what** format will work best.

The following **Report Planning Work Box** is filled out for two separate reports for the same implementation evaluation. Another blank form for your use follows this section.

◊ **Report Planning Work Box**

Type of Report	To Whom	How the report will be used.	Possible Best Format
Verbal	*Education program instructors and other staff.*	*To brainstorm on program changes.*	*Informal interactive workshop*
Written	*Program director.*	*As documentation of program implementation to be included in annual report to grant agency.*	*Detailed type written report that matches description of program plan in original application point by point.*

Evaluating the Evaluations: End-of-Program Evaluation Forms

Perhaps the most common way worker education programs are evaluated is by using forms distributed to students at the end of a training day or program. Commonly the forms ask students to rate different aspects of the programs on a scale of 1 to 5 coupled with a few open-ended questions. Topics on these forms might include the performance of instructors or outside speakers, the training materials such as a manual or video, the style of training used, the usefulness of the program, and so on.

The forms often serve important purposes for monitoring how the program is going and for making improvements or changes. Sometimes, however, those on the front lines, students and instructors, feel that the evaluation forms are of little value, that the information on them isn't really used or that their primary purpose is to fulfill someone else's bureaucratic needs. One health and safety education staff person expressed his frustration this way,

> *As well as I can tell nobody does anything with the information. I don't think they serve any purpose. We don't do anything with them. They give the impression of democracy, but it serves more of a propaganda purpose. If you're not going to put the information to good use, don't collect it.*

Similarly, a worker instructor commented,

> *We keep asking students what they would like changed in our manual. We're not changing the manual. Why not ask different questions - something useful.*

Even if the forms once had a valuable purpose, their routine use may make both the process and the information gathered seem stale and a waste of time. Often the way the information will be used hasn't been fully explored fully.

Improving End of Program Evaluations

As with any routine evaluation or monitoring system, students, instructors and other staff should now and then be asked:

- **How could we improve these evaluation forms** so that they would be more useful?
- **How could we make the questions more relevant** to the current needs of the program, its students and its instructors?
- **How could we use open ended questions** to help learn about unexpected reactions to the program or to gain unanticipated information about how the program is working?
- **How could we better use the information** we gather?
- **How do we let students the information will have an impact** on the program?
- **What other information or methods of obtaining it could we use** instead of, or along with, end-of-the-program evaluation forms?
- **How can we piggy-back gathering of other key bits of information** onto these forms, for example, by asking questions about students' perceptions of the degree of hazards at their work sites, or about other health and safety training students have received.

If end-of-the-program evaluation forms are to serve the program, those on the front lines need to be involved in an ongoing review of both the forms and their use. This review should include instructors, students and others involved in the program. Each of these parties can play a role in ensuring that the information gathered is useful for understanding and improving the program. Time, resources and energy for evaluation are almost always very limited. It must be spent well.

Some alternatives to end-of-the program evaluation forms

There is often concern that ratings on end-of-the-program evaluation forms are inflated, that workers filling them out appreciate the efforts of those responsible for the program and that they want to be supportive of the program. To get more useful information a number of programs have used some different methods.

Example 1: Closed door sessions led by worker-students for worker-students.

Each class selected a student representative responsible for gathering and reporting feedback on the program from fellow students. At the end of the program this person was in charge of leading a confidential, closed-door, student-only discussion about the program. The use of open-ended questions gave students the freedom to express the full range of their reactions to the program. The discussions covered topics similar to an end-of-the-program evaluation form, but also were open to feedback on other topics. The student representative was responsible for taking notes and reporting students' responses (without names) to instructors and staff. Using this format many concerns were voiced and changes were made that otherwise might not have been.

Example 2: Workshop evaluations.

Others programs reserve the final session of the program for an instructor-led open discussion with the entire class. The discussion can cover many of the same topics as a typical end-of-the-program evaluation form or it can be more of a brainstorming session. One worker trainer uses a general question as the basis for discussion: "How can we improve the program so you can go forward and do something with the information and skills you've gained?" Notes are taken on flip chart sheet which may later be typed up and used when program planners work on program changes or when they plan new programs.

Resources for the Evaluation

The following **Evaluation Resource Work Box** is to help you think about what resources you will need for the implementation evaluation. Think of the activities you have discussed, meetings you will need to hold, information you will need to gather, the time it will for various activities, travel expenses and so on. For each activity think about the personal, financial and other resources that will be needed.

◊ **Evaluation Resource Work Box**	
Resource	**Description**
• Personnel _____	
_____	
_____	
_____	
• Budget	
♦ Personnel	
♦ Consultants	
♦ Equipment	
♦ Phone/Supplies	
♦ Travel	
♦ Other	
♦ **Total**	
• Other resources	

8. Evaluating Program Effects: Impacts and Outcomes

Worker health and safety education programs are created to provide workers with tools to advocate or make changes to prevent injuries and illness.

Evaluating program effects seek to find out whether the program:

- **has been effective** in bringing about the intended or desired impacts or outcomes,
- **is worthwhile**, should be continued, expanded, ended or replicated.[32, 33]

The aim of these evaluations is to better understand what difference the program made. Often this difference is thought of in comparison to: a) *not* having participated in the program, or b) having participated in a different program. Sometimes evaluations don't attempt to make comparisons but simply attempt to document changes that have occurred over time for those who have participated.

Evaluations of program effects are often looked at in terms of:

- **changes *expected to lead to* improved health and safety, and**
- ***actual improvements* in the rates of injuries and illness.**

This section describes issues related to evaluating the results of a program.

THE JARGON: Evaluations that look at the effects from a program are called **Summative Evaluations.** They include:

- **Impact Evaluations** - collecting and using information to understand the effects a program had on knowledge, skills, beliefs, attitudes, actions or other characteristics which are believed to affect worker health and safety. Impacts can be measured on the individual, group or organizational level. Impact evaluations can also be called evaluations of intermediate outcomes.

- **Outcome Evaluations** - collecting and using information to understand the effects a program had on the prevention of injuries and illness.

(Process Evaluations are also sometimes included under the category of Summative Evaluations. In this chapter we included them in Section 7, Evaluating Implementation: What Happened in the Program)

Evaluations of Changes Expected to Lead to Improved Health and Safety

The effects education programs have on health and safety, injuries and illness, are not direct. Whatever effects programs have occur through **changes in**:

- **workers' capabilities,** such as knowledge, and skills
- **other factors** that affect how and why workers act, such as
  - confidence
  - attitudes
  - support from others
  - beliefs
  - norms
  - union or organizational support
- **workers' actions**, or
- **conditions** affected by workers' actions

Programs and evaluations may be designed to focus on changes in:

- **individuals**
- **groups**
- **organizations**
- **communities**

Programs and evaluations may also target improvements in:

- **personal protective actions** of individuals, or groups, or
- **social actions** aimed at changing organizations, work site programs, policies or conditions, or public policy and regulations.

The **purpose of an impact evaluation** is to find out to what extent a program leads to changes in *those factors believed to affect* the rate of worker injuries and illness. Impact evaluations can examine any of these various dimensions.

Evaluations of a Program's Effects on Injuries and Illness

- While worker health and safety education programs aim to prevent injuries and illness, measuring changes in injury and illness rates would be an ideal measure of program effectiveness, this is rarely the type of evaluation a health and safety education program can undertake. The reasons are many. Good, available data on injuries and illness there often are hard to obtain. In addition, the rates of injuries and illness are controlled by a number of factors making isolating the effects of education very difficult.

As a result, other types of program results may be targeted as "good indicators" of prevented injuries and illness and be the focus of an *outcome evaluation*. Indicators could include:

- measured reductions in exposures to hazardous conditions

  Changes in levels of exposure to lead.

- Engineering or administrative changes that eliminate the possibility of exposures.

  Substitution of less toxic chemicals for highly toxic ones, installation of process enclosures, or ending the practice of workers entering confined spaces.

While the **purpose of an outcome evaluation** may be to show the extent to which a program has been effective in preventing illness and injuries, this purpose is often seen as establishing the likelihood that the program was a "cause" of these effects.[34] From the viewpoint of "scientific evaluation," establishing that a program "caused" changes is not realistic for most health and safety education programs (see p. __ in Section 4 of this chapter for a discussion of cause).

Some evaluators group these indicators of prevented injuries and illness under the category of ***impact evaluations***. The category isn't all that important. However, to avoid complication, this section will use the term *impact evaluation* for these types indicators.

Integrating Process and Impact Evaluations

While this chapter makes distinctions between *process evaluations*, *how* a program had its effects, and *impact evaluations* designed to measure those effects, there are large areas of overlap. If both types of evaluation are going to be done it is best that *process* and *impact evaluations* be designed together. It is more efficient to gather information for both purposes at the same time.

Selecting Impact Evaluation Questions

Section 6 of this chapter discussed the pluses and minuses of using goals and objectives as a basis for an impact evaluation and presents a step by step basis for writing goals and objectives with a focus on evaluation. If you choose to make goals and objectives the basis for an impact evaluation you should refer to that section. Goals and objectives are not, however, the only or best basis for conducting an impact

evaluation. An alternative is presented here that focuses more directly on the types of intended program impacts that the evaluation team decides are most important.

Linking Levels of Program Action: A Basis For Evaluation Questions[35,36]

Most often our programs are designed believing that **learning** will lead to **actions**, and actions to **change**. Each of these factors, learning, actions and change, can be thought of in the **short- and long-term**. In addition, effects can be thought of and measured on the level of:

- **individuals** - students or others they affect,
- **groups** - the health and safety committee or an emergency response or confined space entry rescue team,
- **organizations** - the local union or the company, or
- **communities** - some programs, like chemical accident prevention, or lead or asbestos abatement in housing involve members of the general community.

Learning or actions that lead to change on one level may lead to learning, actions and change on other levels.

For example, an education program may promote changes in knowledge and skills on the individual level **which lead to**

- improved skills for a chemical emergency response team, **which contributes to**
- the team taking action to affect the organization and procedures used in chemical emergency response, **which leads to**
- the team seeking the help of the health and safety committee **which leads to**
- getting the company to make changes in policies and programs related to chemical accident prevention, and response, **which leads to**
- involvement of municipal firefighters, EMTs and community groups.

Sorting Out and Setting Priorities: Looking for the Most Important Effects

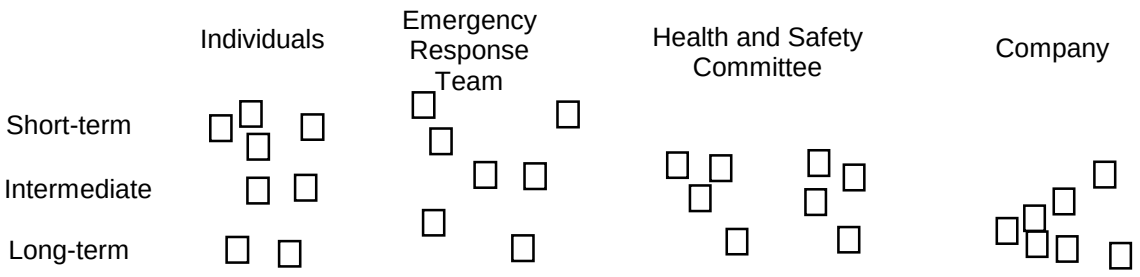
When trying to determine which expected impacts or outcomes are most important for an evaluation, some sorting out and prioritizing is needed. The following **five step process** is designed as a workshop exercise for evaluators, the evaluation advisory

team and others to help sort through and prioritize the factors and levels that could be the target of the evaluation.

1. **Brainstorm** possible program impacts that you might want to measure with members of the evaluation advisory team or others. When doing this think of the short-term, intermediate and long-term effects. Also think about the various levels on which the program might have effects, that is, individual, group, organizational and, if applicable, community. As ideas about effects that could be measured come up, write each one on a separate piece of paper and tape them to wall. You could also write each one on a sheet of one of those sticky-backed notes and arrange these on a wall or sheet of flip chart paper. Make sure everyone can read them. To the extent possible group the ideas in the categories listed above and merge those that have a great deal of overlap.

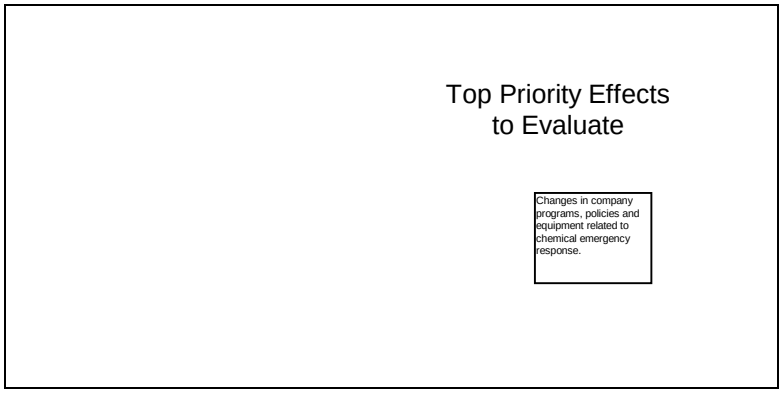
Your wall might look like this.

Brainstorming ideas of impacts to measure.



2. **Focus in** on possible intermediate to long-term range impacts first.
   - As a group select one effect that is among the highest of priorities for the evaluation.
   - Think about whether it is realistic that this kind of effect could occur and be measured easily with the time frame you'll have to do the evaluation.
   - Also keep in mind all the interests that need to be served. For example, the funding agency may require that certain types of outcomes be reported.
   - Move the sheet that contains the effect you selected to the far right-hand side of a second wall or flip chart sheet.
   - Stop and do some more preliminary thinking about how you would measure this kinds of effect. Ask yourselves if it's realistic to expect this kind of effect to be found among

enough individuals, groups, sites or organizations. If at all possible, build consensus. Your second wall would now look like this.



3. **Effects Leading to Other Effects.** Now, thinking about the effect that you selected and isolated, look again at the original grouping of impacts. Which of these are most important in leading to the impacts you have already selected and placed on the second wall? This may take some juggling, moving sheets back and forth and some intense, and perhaps lengthy discussion. Once decided upon, this would be an intermediate effect you might also want to look at in the evaluation.

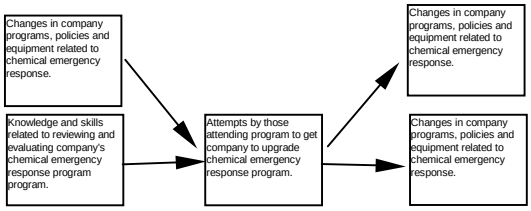
4. **Short-term Effects Leading to Intermediate Effects.** Repeat step 3, except this time look for an item you believe is most important for influencing the intermediate effect you identified. In this step it is more likely that you are identifying a short-term effect of the program, for example a change that can be examined right after the program.
5. **Draw arrows** to show how you believe each element is affecting others. When completed record the diagram on a piece of paper that can be used as a permanent record.

Any of these steps may be repeated to add more effects and detail. If you choose to repeat these steps, keep thinking about the time and resources that will be needed to conduct an evaluation of your collection of effects. If the diagram gets too complex, scale it back or decide only to consider a more limited set of effects in the evaluation.

Keep in mind that very short-term effects may be easier to measure but give you less useful information about whether you have achieved your longer-term aims. Longer-term effects may be harder to measure and it may be more difficult to sort out whether any measured effects were a result of the education program or other factors. However, these longer-term effects are likely to be closer to the goals you actually trying to achieve. It may be easy for a program to get stuck focusing in on short-term

influences such as knowledge and skills, rather than focus on important underlying causes of injuries and illness such as poor work design, health and safety programs, policies, procedures, equipment, or work organization.

Short-term and intermediate effects that lead to top priority effect to evaluate.



Having information about program impacts on these different levels can give you a richer understanding of your program and its effects. It can tell you the kinds of things you can and cannot expect to be affected by your program. It can also help inform you about which short-term outcomes are most important for achieving long-term outcomes. Coupled with information from an implementation evaluation, information on these various levels begins to help make connections between the activities and outcomes.

Key Impact Evaluation Questions and Information Needed to Answer Them

Once the evaluators have decided on key objectives or other indicators for the impact evaluation, the questions that will guide the evaluation should be written. Questions using a quantitative approach will be trying to answer questions about how much or how many. Often evaluation questions focusing on a *quantitative approach* can be written starting with the phrase: *To what extent ....* The question asks to what extent something has happened or changed. Questions using a *qualitative approach* will be trying to answer how, what and why questions. Q*ualitative* methods help uncover greater detail about specific cases. The evaluation may focus on either quantitative or qualitative approaches, or it may use both.

***A quantitative evaluation question**: To what extent have grain mill workers who attended the two-day confined-space entry education program attempted to engage the company in a review process of the existing confined space entry program.*

***A qualitative question**: For those local unions that report the occurrence of confined space entry incidents or near misses following the program, what happened in the incidents and how and why did the education program have or not have an impact on the incidents.*

Again, think about time, energy and resources it will take to answer these more specific questions. Keep asking yourselves whether the scope of your evaluation focus is reasonable. Once the evaluation team has agreed upon the evaluation questions to be studied, the evaluation plans can begin to focus on methods and sources for data collection.

Using your goals and objectives or other ways in which you have identified impacts you want to measure, write specific evaluation questions using the **Impact Evaluation Question Work Box**.

◊ **Impact Evaluation Question Work Box**

Write one evaluation question for each major impact and type of method you will use?

⇒

⇒

⇒

The findings from an impact evaluation may have serious consequences for a program. Decisions on whether to continue a program, or to increase or decrease funding may depend in some part on the results of the evaluation. **Look at each of the evaluation questions and consider the possible consequences of either positive or negative findings.** How would various interested parties react? Its best to consider and weigh these possibilities before you begin.

The Design of the Evaluation

Once evaluation questions are set evaluators must decide which design or to use. Issues related to evaluation design are covered in sections of this chapter on qualitative and quantitative approaches to evaluation. The design you choose should the one that is most compatible with the evaluation questions, time energy and resources and the knowledge and abilities of those conducting the evaluation. Evaluation designs may include:

- **single group studies** using either a before and after or one-time approach,
- **comparison group studies**,
- **goals and objectives studies** focusing on whether or not targets have been achieved, and
- **case studies** - of individuals, groups or organizations which look at either single or multiple cases.

The following is a **Impact Evaluation Study Design Work Box** for you to record the study design the evaluation will use and the persons, groups or organizations for which the data is analyzed. A sample has been completed for the confined space program used earlier in this section. Another blank form for your use follows this section.

◊ Impact Evaluation Study Design Work Box

Briefly describe the evaluation design that will be used and the persons, groups or organizations for which the data is analyzed.

⇒ **Design:** *Case study of four sites.*

⇒ **Persons, groups or organizations for which the data will be analyzed:**

The evaluation will compare and contrast sites where workers have attended the confined space education program and that have had confined space incidents or near incidents in the six months following the program.

Choosing Data Collection Sources and Methods for the Evaluation

As with implementation evaluations, impact evaluations may use a variety of sources and methods to gather information. For a discussion of these issues refer the Choosing Data Collection Sources and Methods for the Evaluation sub-section in the Evaluating Implementation section of this chapter. That section contains a sample **Sources and Methods of Information Collection Work Box** which is also relevant to impact evaluations. A blank Work Box for your use follows this section.

Analysis of and Use of Information[37]

The design and methods used for the evaluation will dictate how the information collected should be analyzed. Scores, averages or other statistical measures should be computed from quantitative data sources. Major ideas or themes should be identified and organized from qualitative data (see previous sections). Drawings, graphs or charts can be used if they help clarify results and are an effective way of communicating for those receiving the report. The major findings and the supporting evidence should be organized and developed into an oral and/or a written presentation.

Just as with and implementation evaluation, keep in mind:

- **who needs to be informed about the evaluation process, its purposes and activities**, and
- **how will they be informed, kept up to date and involved in the process.**

Different groups that may be interested in the evaluation findings may have different needs and interests in terms of the presentation of the data.

The Meaning of Information Collected

Meaning is what one makes of the information that is collected to answer the evaluation questions. In an impact evaluation this phase is used to examine whether the program:

1. **has been effective**,
2. **is worthwhile**.

If the evaluation has examined more than one level of program action the possible connections between the findings of these levels can be examined. Be careful not to overstep the limits of the design and methods you have used.

When the impact evaluation is coupled with a process evaluation, not only can the evaluations shed light on changes that may have occurred, but they can help us understand how and why changes did or did not occur.

Impact Evaluation Reports

Different groups who have a desire and need to be informed about the findings of the evaluation may have different needs in terms of how the evaluation is reported. Often a more formal, detailed report can be written for a funding agency or program sponsors and a summary developed for others who would be less interested in the details of the report. Typically a formal report should include:

- **an introduction** that briefly describes the program, its goals and methods
- a description of the evaluation **design and the methods** used to collect and analyze information,
- **a description of who participated** in the program and the evaluation including how people were selected,
- **results** of the evaluation, and
- **conclusions** from the results.

Just as with an implementation evaluation, the ultimate usefulness of an evaluation of program impacts depends not just on the evaluators, but also on those who implemented the program or were affected by it. Analysis and interpretations of findings will be enriched if representatives from a group like an evaluation advisory team have an opportunity to contribute. To serve various interests reports should be made in three steps:

- The initial report **should be created by the evaluators and -**
- shared and discussed **with the evaluation team which works towards agreement on modifications to the -**
- final report **to be made by the evaluators.**

In writing the reports **think about:**

- who **will use the reports,**
- when **they will need them,**
- how **the reports will be used, and**
- what **format will work best.**

The **Report Planning Work Box** is filled out for two separate reports for the same impact evaluation. Another blank form for your use follows this chapter.

◊ **Report Planning Work Box**

Type of Report	To Whom	How the report will be used.	Possible Best Format
Written summary	*Education program instructors and other staff and participants.*	*To inform and celebrate any successes.*	*Brief, user friendly written summary of detailed report.*
Written detailed report	*Funding agency and program director.*	*As documentation of program impacts to be included in annual report to grant agency and upcoming grant proposals.*	*Detailed type written report that matches statement of goals and objectives in the original application.*

1 Patton, M. Q., 1986. Utilization Focused Evaluation (2nd ed.). Newbury Park, CA: Sage Publications.

2 Steckler, A., McLeroy, K. R., Goodman, R. M., Bird, S. T. and McCormick, L. 1992. Toward integrating qualitative and quantitative methods: An introduction, Health Education Quarterly, 19(1), 1-8.

3 Clark, N. M. and McLeroy, K. R. 1995. Creating capacity through health education: What we know and what we don't, Health Education Quarterly, 22(3), 273-289.

4 Steckler et al. 1992.

5 Patton, M. Q., 1986.

6 Fitz-Gibbon, C. T. and Morris, L. L. 1987. How to Design a Program Evaluation (Volume 3 8 of a 9 volume series on program evaluation). Newbury Park, CA: Sage Publications.

7 McQuiston, T. H., Coleman, P., Wallerstein, N. B., Marcus, A. C., Morawetz, J. S. and Ortlieb, D. W. 1994. Hazardous Waste Worker Education: Long term effects. Journal of Occupational Medicine, 36(12): 1311-1323.

8 Marcus, A. C., Baker, D. B., Froines, J. R., Brown, E. R., McQuiston, T. and Herman, N. A. 1986. ICWU Cancer Control Education and Evaluation Program: Research design and needs assessment. Journal of Occupational Medicine, 28 (3), 226-236.

9 Patton, M. Q., 1986.

10 Judd, C. M., Smith, E. R. & Kidder, L. H. (1991). (6th Edition). Research Methods in Social Relations. Holt, Rinehart & Winston, Inc. pp. 213-265.)

11 Singleton, R. A., Straits, B. C. and Straits, M. M. 1993. Approaches to Social Research (2nd ed.). New York: Oxford University Press.

12 Judd, et al. 1991.

13 Singleton, et al. 1993.

14 Adapted from Table 10.1 Summary Table of Different Data collection Methods. p 223. Judd, C. M., Smith, E. R. and Kidder, L. H. 1991. Research Methods in Social Relations (6th ed.). Fort Worth, TX: Holt, Rinehart and Winston Inc.)

15 Fitz-Gibbon, C. T. and Morris, L. L. 1987. How to Analyze Data (Volume 8 of a 9 volume series on program evaluation). Newbury Park, CA: Sage Publications.

16 Wisconsin Department of Human Services. 1981. Four steps to better objectives.

17 Kettner, P. M., Moroney, R. M. and Lawrence, L. M. 1990. Designing and managing programs: An effectiveness-based approach. Sage Publications: Newbury Park, CA.

18 McQuiston, et al, 1994.

19 Parker-Brown, M and Nguyen-Scott, N. 1992. Evaluating a training-for-action job health and safety program. American Journal of Industrial Medicine, 22: 739-749.

20 King, J. A., Morris, L. L., Fitz-Gibbon, C.T. 1987. How to Assess Program Implementation (Volume 5 of the 9 volume Program Evaluation Kit). Newbury Park, CA: Sage Publications.

21 Dehar, M. A., Casswell, C. and Duignan, P. 1993. Formative and process evaluation of health promotion and disease prevention programs. Evaluation Research, 17: 204-220.

22 Goodman, R. M. & Wandersman, A. 1994. FORECAST: A formative approach to evaluating community coalitions and community-based initiatives. Journal of Community Psychology, Special Issue, 6-25.

23 Dehar, et al. 1993.

24 Israel, B., Cummings, M., Dignan, H., Heaney, C., Perales, D. Simmons-Morton, B. and Zimmerman, M. 1995. Evaluation of health education programs: Current assessment and future directions, Health Education Quarterly, 22: 364-389.

25 Chen, H. and Rossi, P. H. 1983. Evaluating with sense: The theory driven approach. Evaluation Review, 7, 283-302.

26 Patton, M. Q. 1986.

27 King, et al, 1987.

28 Patton, 1986.

29 Patton, 1986.

30 Herman, J. L., Morris, L. L. and Fitz-Gibbon, C. T. 1987. Evaluator's Handbook. (Volume 1 of the 9 volume Program Evaluation Kit Newbury Park, CA: Sage Publications.

31 Goodman, & Wandersman. 1994.

32 Rossi, P. H. & Freeman, H. E. 1993. Evaluation: A systematic approach (5th ed.). Sage Publications: Newbury Park, CA.

33 Patton, M. Q. 1986.

34 Rossi, P. H. & Freeman, H. E. 1993. Evaluation: A systematic approach (5th ed.). Sage Publications: Newbury Park, CA.

35 Harris, E. M. 1992. Assessing community development Research methodologies, Canadian Journal of Public Health, Supplement 1, S62-66.

36 Lipsy, M. W. and Pollard, J. A. 1989. Driving toward theory in program evaluation: More models to choose from. Evaluation and Program Planning, 12, 317-328.

37 Herman, et al. 1987.